



# DiRAC déploie la solution Atempo Miria for Archiving

CLIENT : DIRAC/UK

SECTEUR : RECHERCHE AVANCÉE / HPC



Quatre universités britanniques, Cambridge, Durham, Edimbourg et Leicester, assurent la gestion des architectures HPC du DiRAC.

Ces universités fournissent les installations de calcul haute performance principales, ainsi que l'expertise nécessaire à la réalisation de multiples projets de recherche, de soutien, de transfert de connaissances, et de partenariat industriel.

*Les données jouent un rôle important dans la mission du DiRAC; les besoins en puissance de calcul (mesurés en pétaflops) et en stockage de données (en pétaoctets) sont au cœur des projets soutenus par le DiRAC.*

## À propos du DiRAC

Le DiRAC (Distributed Research Utilizing Advanced Computing), recherche distribuée basée sur les calculs haute performance) est le centre de calcul intensif intégré pour la modélisation théorique et la recherche basée sur le calcul haute performance (HPC) en physique des particules, astronomie et cosmologie, et un élément-clé du Science and Technology Facilities Council (STFC, Conseil des installations scientifiques et technologiques) du Royaume-Uni. Les chercheurs du monde entier peuvent accéder aux puissantes installations informatiques du DiRAC.

## À propos du DiRAC et d'Atempo

Le centre Memory Intensive du DiRAC à Durham a récemment fait appel à Atempo, le spécialiste de la protection et du transfert de données, et à son partenaire britannique OCF pour mettre en œuvre un projet d'archivage multi-pétaoctets de ses données Lustre et Spectrum Scale (GPFS).

Le projet DiRAC Memory Intensive, septième cluster HPC à Durham, permet de fournir aux chercheurs une puissance de calcul de 452 nœuds et 12 656 cœurs.

Nous avons rencontré le Dr Lydia Heck, Directrice Technique et le Dr Alastair Basden, Responsable Technique du DiRAC qui nous ont donné leur avis sur la solution Miria for Archiving d'Atempo, et comment elle répond à leurs exigences d'archivage des projets de recherche sur bande. Ils ont également évoqué l'évolution possible de leurs besoins en archivage et sauvegarde de données des mois à venir.



Le DiRAC@Durham a eu pour mission récente la transition du système de stockage Spectrum Scale (GPFS) vers le système Lustre de DDN. L'objectif était, selon Lydia, « d'augmenter efficacement la capacité de stockage en utilisant une solution moins coûteuse ». Le besoin croissant de mémoire de calcul génère en effet d'importants volumes de données dans les environnements HPC.

La puissance de calcul supplémentaire est étroitement liée aux besoins accrus en matière de stockage et de hiérarchisation de données. Le stockage des données nécessaire d'ici 2022 devrait atteindre 20 Po avec le déploiement de la phase 3 du DiRAC.

Le plan de gestion des données du DiRAC comprend l'archivage de la base de données de recherche et des documents de recherche scientifique finalisés, révisés par des pairs.

## La solution dans la pratique : Miria for Archiving d'Atempo

Le rôle de Miria est d'archiver les données de recherche et d'alléger le stockage sur disque, plus coûteux, en transférant les données de recherche du stockage principal vers quatre bandes LTO. Les besoins actuels de stockage sont de 6,5 Po et passeront à 20 Po lorsque la phase 3 du DiRAC HPC sera déployée (prévue pour début 2020), ce qui multipliera par 10 la vitesse de traitement et de création de données, et par conséquent des besoins en stockage.

La solution d'archivage en place était lente et pas suffisamment évolutive. Atempo a proposé fin 2018 une Preuve de concept (POC) d'abord sur Spectrum Scale (GPFS) puis sur Lustre.

Atempo a rapidement démontré que Miria for Archiving était une solution idéale pour répondre aux besoins d'archivage de données du DiRAC. Un serveur d'archivage Miria et son Data Mover dédié accèdent directement au système de fichiers Lustre et gèrent efficacement tous les flux d'archivage de données de la source à la destination.

L'interface utilisateur de Miria for Archiving permet aux utilisateurs d'effectuer leurs propres tâches d'archivage rapidement et de restaurer directement leurs données. Les chercheurs peuvent utiliser l'arborescence logique de leur choix pour chaque projet de recherche lors de l'archivage de leurs données. L'interface d'administration permet au DiRAC de gérer l'archivage et la sauvegarde des données critiques.



*« Miria for Archiving est extrêmement puissant et riche en fonctionnalités, et devrait répondre à nos besoins futurs. Nous avons été très impressionnés par Miria, même si nous n'avons entrevu qu'une petite partie de ses possibilités. »*

*Dr Alastair Basden,  
Responsable Technique du  
Centre Memory Intensive  
du DiRAC,  
Université de Durham*



Les équipes du DiRAC utilisent également Miria pour sauvegarder les répertoires personnels des utilisateurs ainsi que des fichiers d'archivage sélectionnés par ces derniers. « Chaque action de l'administrateur et de l'utilisateur est gérée par le protocole HTTP, ce qui signifie que nous pouvons très facilement mettre en place un tunnel SSH et permettre aux utilisateurs d'archiver leurs fichiers où qu'ils soient dans le monde. » Tous les équipements physiques de transfert de données ainsi que le serveur Miria sont installés et opérationnels, et Alastair Basden précise que : « les performances d'archivage sur les flux de données des systèmes de fichiers Lustre sont optimales, ce qui nous satisfait pleinement ».

Les exigences d'archivage imposent la création de copies des données sur deux bandes distinctes. Le DiRAC utilise aujourd'hui des bandes LTO7 et prévoit de passer aux bandes LTO8 dès que celles-ci seront disponibles. Atempo supporte pleinement ce mélange de technologies.

## Perspectives

Miria n'est pas seulement un outil d'archivage. Le composant Miria for Backup assure également la protection des données des utilisateurs. Au cours du POC, Atempo a aussi démontré les capacités de FastScan pour GPFS. FastScan optimise la reconnaissance et le stockage des nouveaux fichiers et des fichiers modifiés afin d'éviter de multiples et fastidieux balayage du système de fichiers. À l'avenir, le DiRAC pourra intégrer les capacités de sauvegarde avancées de Miria avec la future capacité FastScan d'Atempo pour Lustre de DDN.

La force des équipes R&D d'Atempo réside dans leur connaissance approfondie des systèmes de fichiers du marché et dans leur capacité à manipuler les attributs de fichiers et les options de striping. Le DiRAC souhaite collaborer avec les équipes DDN et Atempo pour réaliser cette intégration, et leur donner accès à ses environnements de systèmes de fichiers et de stockage à l'échelle HPC afin de tester la solution FastScan pour Lustre.

« C'est exactement ce que le DiRAC attend de ses fournisseurs de technologie actifs. Pour justifier des dépenses aussi importantes, nous devons pouvoir garantir à nos actionnaires que nous sommes à l'avant-garde de l'innovation dans de nombreux domaines, dont le transfert et la protection de données », conclut Lydia Heck.

Le DiRAC possède plusieurs sites à connexion haut débit - entre les universités membres (bande passante pouvant atteindre 20 GBit/s) et prévoit de pouvoir transférer, protéger et stocker les données en dehors de ses sites. L'organisme envisage par ailleurs la possibilité de faire une copie secondaire des données d'un site et de les stocker sur un autre site.

En conclusion, la solution Miria for Archiving a produit des résultats très positifs en matière de performance, d'efficacité et de sécurité. Les équipes d'Atempo et du DiRAC continueront à travailler dans un esprit de coopération et de mutualisation des ressources, afin de confirmer ce premier succès.





Atempo, éditeur de logiciels français et leader européen avec une présence globale, propose des solutions pour sauvegarder, protéger, stocker, déplacer et restaurer les données critiques de milliers d'entreprises dans le monde.